



Multi-modal and multi-vendor retina image registration

ZHANG LI,^{1,2,7} FAN HUANG,^{3,7} JIONG ZHANG,³
BEHDAD DASHTBOZORG,³ SAMANEH ABBASI-SURESHJANI,³
YUE SUN,⁴ XI LONG,⁴ QIFENG YU,^{1,2}
BART TER HAAR ROMENY,^{3,5} AND TAO TAN^{3,6,*}

¹ College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China

² Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China

³ Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands

⁴ Electrical Engineering, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands

⁵ Department of Biomedical and Information Technology, Northeastern University, Shenyang, 110000, China

⁶ Research and Development, ScreenPoint Medical, Nijmegen, 6512 AB, The Netherlands

⁷ These authors contributed equally to the work

*tao.tan911@gmail.com

Abstract: Multi-modal retinal image registration is often required to utilize the complementary information from different retinal imaging modalities. However, a robust and accurate registration is still a challenge due to the modality-varied resolution, contrast, and luminosity. In this paper, a two step registration method is proposed to address this problem. *Descriptor matching* on mean phase images is used to *globally* register images in the first step. *Deformable registration* based on modality independent neighbourhood descriptor (MIND) method is followed to *locally* refine the registration result in the second step. The proposed method is extensively evaluated on color fundus images and scanning laser ophthalmoscope (SLO) images. Both qualitative and quantitative tests demonstrate improved registration using the proposed method compared to the state-of-the-art. The proposed method produces significantly and substantially larger mean Dice coefficients compared to other methods ($p < 0.001$). It may facilitate the measurement of corresponding features from different retinal images, which can aid in assessing certain retinal diseases.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (100.2960) Image analysis; (100.0100) Image processing.

References and links

1. N. Amerasinghe, T. Aung, N. Cheung, C. W. Fong, J. J. Wang, P. Mitchell, S.-M. Saw, and T. Y. Wong, "Evidence of retinal vascular narrowing in glaucomatous eyes in an asian population," *Invest. Ophthalmol. Vis. Sci.* **49**(12), 5397–5402 (2008).
2. L. S. Lim, P. Mitchell, J. M. Seddon, F. G. Holz, and T. Y. Wong, "Age-related macular degeneration," *Lancet* **379**(9827), 1728–1738, (2012).
3. T. Y. Wong, N. Cheung, W. T. Tay, J. J. Wang, T. Aung, S. M. Saw, S. C. Lim, E. S. Tai, and P. Mitchell, "Prevalence and risk factors for diabetic retinopathy: the singapore malay eye study," *Ophthalmology* **115**(11), 1869–1875 (2008).
4. T. MacGillivray, E. Trucco, J. Cameron, B. Dhillon, J. Houston, and E. Van Beek, "Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions," *Br. J. Radiol.* **87**(1040), 20130832 (2014).
5. R. H. Webb, G. W. Hughes, and F. C. Delori, "Confocal scanning laser ophthalmoscope," *Appl. Opt.* **26**(8), 1492–1499 (1987).
6. M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.* **3**, 169–208 (2010).

7. B. I. Gramatikov, "Modern technologies for retinal scanning and imaging: an introduction for the biomedical engineer," *Biomed. Eng. Online* **13**(1), 52 (2014).
8. A. V. Cideciyan, "Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors," *IEEE Eng. Med. Biol. Mag.* **14**(1), 52–58 (1995).
9. R. Kolar, L. Kubecka, and J. Jan, "Registration and fusion of the autofluorescent and infrared retinal images," *Int. J. Biomed. Imaging* **2008**, 513478 (2008).
10. P. A. Legg, P. L. Rosin, D. Marshall, and J. E. Morgan, "Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation," *Comput. Med. Imaging Graph.* **37**(7), 597–606 (2013).
11. N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, "Registration of stereo and temporal images of the retina," *IEEE Trans. Med. Imag.* **18**(5), 404–418, (1999).
12. Z. Yi and S. Soatto, "Nonrigid registration combining global and local statistics," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2009)*, pp. 2200–2207.
13. A. Andronache, M. von Siebenthal, G. Székely, and P. Cattin, "Non-rigid registration of multi-modal images using both mutual information and cross-correlation," *Med. Image Anal.* **12**(1), 3–15 (2008).
14. C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Trans. Med. Imag.* **22**(11), 1379–1394, (2003).
15. J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan, "Optic disc feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recogn.* **40**(7), 2063–2076 (2007).
16. H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Trans. Biomed. Eng.* **51**(2), 246–254 (2004).
17. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110, (2004).
18. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Und.* **110**(3), 346–359 (2008).
19. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2005)* **1**, pp. 886–893.
20. T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987, (2002).
21. M. S. Miri, M. D. Abràmoff, Y. H. Kwon, and M. K. Garvin, "Multimodal registration of SD-OCT volumes and fundus photographs using histograms of oriented gradients," *Biomed. Opt. Express* **7**(12), 5252–5267 (2016).
22. L. Chen, Y. Xiang, Y. Chen, and X. Zhang, "Retinal image registration using bifurcation structures," in *Image Processing (ICIP), 2011 18th IEEE International Conference on (IEEE, 2011)*, pp. 2169–2172.
23. K. Zhang, E. Zhang, J. Li, and G. Chen, "Retinal image automatic registration based on local bifurcation structure," in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on (IEEE, 2016)*, pp. 1418–1422.
24. J. Ma, J. Jiang, J. Chen, C. Liu, and C. Li, "Multimodal retinal image registration using edge map and feature guided Gaussian mixture model," in *Visual Communications and Image Processing (VCIP), (IEEE, 2016)*, pp. 1–4.
25. M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, "Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Med. Image Anal.* **16**(7), 1423–1435 (2012).
26. M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.* **49**(12), 3136–3144 (2001).
27. A. Cifor, L. Rissler, D. Chung, E. M. Anderson, and J. A. Schnabel, "Hybrid feature-based diffeomorphic registration for tumor tracking in 2-d liver ultrasound images," *IEEE Trans. Med. Imag.* **32**(9), 1647–1656 (2013).
28. A. Wong, D. A. Clausi, and P. Fieguth, "CPOL: Complex phase order likelihood as a similarity measure for MR–CT registration," *Med. Image Anal.* **14**(1), 50–57 (2010).
29. G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)* (The MIT Press, 2006).
30. T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 500–513 (2011).
31. A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 412–419 (2002).
32. P. Cattin, H. Bay, L. Van Gool, and G. Székely, "Retina mosaicing using local features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer, Berlin, Heidelberg, (2006))* pp. 185–192.
33. M. Golabbakhsh and H. Rabbani, "Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model," *IET Image Process.* **7**(8), 768–776 (2013).
34. A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.* **32**(7), 1153–1190 (2013).
35. Z. Li, D. Mahapatra, J. A. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos, "Image registration based on autocorrelation of local structure," *IEEE Trans. Med. Imag.* **35**(1), 63–75, (2016).
36. J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imag.* **35**(12), 2631–2644 (2016).

37. S. Niu, Q. Chen, H. Shen, L. de Sisternes, and D.L. Rubin, "Registration of SD-OCT en-face images with color fundus photographs based on local patch matching," In Proceedings of the Ophthalmic Medical Image Analysis First International Workshop (OMIA, 2014), pp. 25–32.
38. Y. Li, G. Gregori, R.W. Knighton, B.J. Lujan, and P.J. Rosenfeld, "Registration of OCT fundus images with color fundus photographs based on blood vessel ridges," *Opt. Express* **19**(1), 7–16 (2011).
39. E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2007), pp. 1–8.
40. R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A. M. Wood, M. S. Ludwig, and T. Guerrero, "A reference dataset for deformable image registration spatial accuracy evaluation using the COPD gene study archive," *Phys. Med. Biol.* **58**(9), 2861 (2013).
41. K. Adal, P. van Etten, J. P. Martinez, K. Rouwen, L. J. van Vliet, and K. A. Vermeer, "Automated detection and classification of longitudinal retinal changes due to microaneurysms for diabetic retinopathy screening," *Invest. Ophthalmol. Vis. Sci.* **57**(12), 3403–3403 (2016).

1. Introduction

Retinal images reveal biological information about the retina, which are examined by ophthalmologists to diagnose and monitor the progression of a variety of diseases, including diabetic retinopathy, age-related macular degeneration, and glaucoma [1–3]. Retinal images are often acquired by different imaging modalities in order to have multiple representations for the eye [4]. For example, color fundus photography and Scanning Laser Ophthalmoscope (SLO) [5] are two commonly used techniques for retinal image acquisition in ophthalmology [6], where the former one uses the imaging light of red, green and blue wavebands and the latter one uses single wavelength laser light (see Fig. 1 (a) and (b)). Therefore, the retinal blood vessels on SLO images have higher contrast to the background than the ones on color fundus images, while the color images have a larger field-of-view and better discrimination between artery and vein [7]. In a screening setting, ophthalmologists or computer-aided diagnosis system may also evaluate exams from different screen rounds. In the clinical review of disease progression, longitudinal retinal image registration is crucial which establishes a pixel-to-pixel correspondence among the images of different modalities. It provides extra assistance for observers to find the early and subtle signs of abnormalities on one type of images and confirm their findings in the same area of other images.

To make temporal analysis possible or to investigate the same findings from different retinal images, an image registration technique is necessary. However multi-modal retinal image registration is still a challenge due to the modality-varied resolution, contrast and luminosity between different modality images (i.e. color and SLO retinal images). In the last few years, registration approaches proposed in the literature for retinal images can be summarized into two categories: area-based methods and feature-based method. An area-based method aims to extract intensity-based features from the overlapped area of two images, then optimizes some similarity measurements such as cross correlation [8,9], mutual information [10–12] and/or the combination of both of them [13] to obtain the best alignment. A feature-based approach tries to extract descriptive features for finding the correspondence between the images. Commonly-used descriptors include the position of retinal landmarks like vascular bifurcation points [14], the optic disc center [15], and the fovea center [16]. Moreover, high-level feature descriptors such as scale invariant feature transformation (SIFT) [17], speeded-up robust features (SURF) [18], histograms of oriented gradients (HOG) [19] and local binary patterns (LBP) [20] are proposed for registration. An objective function is used based on the correspondence between the extracted feature descriptors to find the best transformation parameters. One of the recent methods [21] was proposed to detect landmarks (corner points) as a preprocessing step. Afterwards, HOG was calculated from the neighborhood of each corner point. During the registration, random sample consensus (RANSAC) was used to remove the incorrect correspondences to achieve the best affine transformation between images.

There are several limitations when using the previous techniques for the registration of color

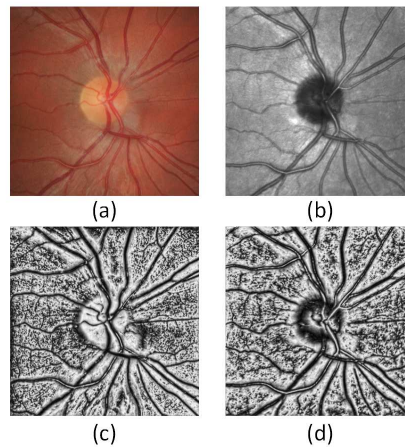


Fig. 1. The retina of the same subject acquired by (a) color fundus camera (Canon Cr-1 Mark II) and (b) Scanning Laser Ophthalmoscope (Spectralis HRA OCT). Retinal landmarks such as blood vessel and the optic disc have different representations on both modalities. (c) and (d) are mean phase images derived from (a) and (b), respectively.

images and SLO images. First of all, the representation of retinal landmarks on both types of image is different. The optic disc is bright on color images but dark on SLO image. On color images, the central reflection of blood vessels is visible on arteries rather than on veins, while on SLO images it can be clearly seen on both types of vessels. Therefore, area-based methods which rely on measuring the similarity between two images are difficult because the similarity metric is not easy to define. In addition, tiny blood vessels on SLO images are depicted more clearly than those on RGB images and the vascular structures on both images might have a quite large differences. Thus feature-based methods which use vascular bifurcation points [22,23] and vessel edge map [24] are not ideally applicable.

In this paper, we propose a new framework for multi-modal retina image registration (specifically of color and SLO images), which addresses the issues of aligning images of different modalities. The methodology consists of two main steps: 1) *descriptor matching* and 2) *deformable image registration*. Descriptor matching is used on mean phase images to estimate the global transformation (e.g affine) between images. The descriptor is densely calculated on mean phase images so that the matching step is invariant to contrast differences. The modality independent neighborhood descriptor (MIND) based deformable registration method proposed by Heinrich *et al.* [25] is then followed to refine the registration result *locally* as the deformable registration could achieve better estimation of the transformation than the affine registration.

2. Method

2.1. Descriptor matching

2.1.1. Modality-invariant feature descriptors

The transformation parameters for image registration are derived by matching the feature descriptors extracted from the images. In order to address the effect of contrast and illumination difference between color and SLO images, we extract descriptors from mean phase images, which are independent on pixel intensity [26]. Firstly, an image $f(\mathbf{x})$ is converted to a phase image $\varphi(\mathbf{x})$ [26] via the followed transformation:

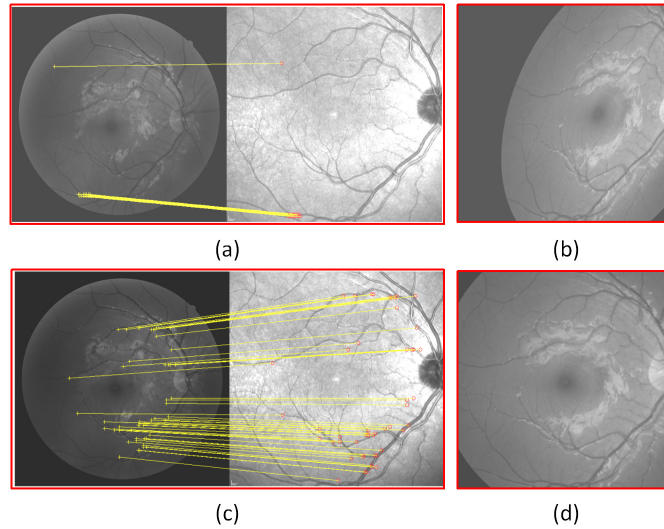


Fig. 2. Descriptor matching result between a color image (moving image) and an SLO image (fixed image, e.g. acquired by a Spectralis HRA OCT camera). (a) and (c) show matching pairs based on HOG and HOG-MP with the RANSAC process; (b) and (d) are the matching result using an affine transformation based on HOG and HOG-MP, respectively.

$$\varphi(\mathbf{x}) = \text{atan}\left(\frac{|\mathbf{f}_R(\mathbf{x})|}{f_e(\mathbf{x})}\right), \quad (1)$$

where $\mathbf{f}_R(\mathbf{x})$ is the Riesz transform proposed by Felsberg *et al.* [26] of $f(\mathbf{x})$ that represents the *odd* component of $f(\mathbf{x})$, and $f_e(\mathbf{x})$ is the *even* component of $f(\mathbf{x})$ [27]. In the proposed framework, $\mathbf{f}_R(\mathbf{x})$ and $f_e(\mathbf{x})$ are measured by a set of log-Gabor filters [28], which have zero DC component with tunable bandwidth in the Fourier domain so that the filter responses are independent of image intensity. In addition, instead of calculating local phase at a certain empirical scale, we use log-Gabor filters with multiple scales σ to derive $\varphi_\sigma(\mathbf{x})$ [27]. Then we compute the average of the local phase images, named mean phase image $\bar{\varphi}(\mathbf{x})$, for the feature descriptors extraction. $\bar{\varphi}(\mathbf{x})$ averages phases over all scales, and serves as an identifier. For example, a step corresponds to $\varphi = 0$ and a peak to $\varphi = \frac{\pi}{2}$. Essentially, it provides contrast-invariant measurement for descriptor extraction (see Fig. 1 (c) and (d)).

On the mean phase images, we densely compute the histogram of oriented gradients (HOG-MP) as the feature descriptors for image matching [19]. A square block of size $MN \times MN$ around each selected pixel is drawn for gradient calculation. The distance between each selected points is 5 pixels in all dimensions. Every block contains $M \times M$ cells which have the size of $N \times N$. For the pixels in each cell, the histogram of the gradient in twelve directions (from 0 to 360 with step size 30) is obtained to extract the local structural information inside the block, which yields $12 \times N \times N$ feature descriptors for each selected point. We choose $M=11$ and $N=3$ in our experiments. Miri *et al.* [21] applied the HOG calculation on detected interesting points from the original intensity images. However, the HOG-based descriptor matching step may suffer from the large contrast differences between multi-modal images. The detection accuracy of the interesting points may also influence the matching results.

2.1.2. Matching method

We first resize the moving images into the same size of fixed image (see Fig.2(a) and (c)). One can see that the size of the same vessel and disc from different images are not the same. Since the HOG-

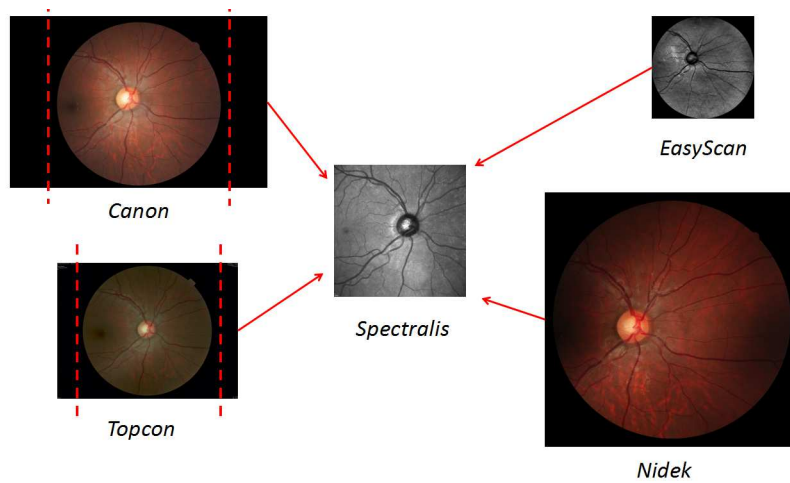


Fig. 3. Five types of eye images need to be registered. The images of Spectralis fundus camera are the fixed image with size of 1536×1536 pixels; images from Canon, Topcon, EasyScan and Nidek are moving images (registered to Spectralis) with the size of 3456×2403 , 2408×1536 , 1024×1024 and 3744×3744 pixels, respectively. The images are shown in relative pixel size.

MP is densely calculated, we use the approximate nearest neighbor search method [29] to efficiently match descriptors. Similar to [30], this method is used twice (searching from the fixed image to the moving image and vice versa) to exclude outliers. Only the correspondences that exist in the two matching results are left. We used their implementation and parameter settings for approximate nearest neighbor search method (see <https://people.eecs.berkeley.edu/katef/LDOF.html>). Then random sample consensus is applied to further remove the incorrect correspondences. An affine transformation is estimated and updated during the RANSAC process [21]. The final affine transformation is then applied to the fundus images.

The descriptor matching results based on HOG-MP and HOG are compared in Fig. 2. It can be observed that the HOG-MP based matching shows good correspondence between color and SLO images, especially around the positions with structures (e.g. vessel bifurcation and crossing points), as expected.

2.2. Deformable image registration

An affine transformation, as a first-order transformation, is not flexible enough to model the deformation between color images and SLO images. A second order quadratic transformation was proposed to improve the deformation modeling [31–33]. However, local deformations caused by the eye movements and breathing during acquisition can not be modeled by the quadratic transformation effectively. Therefore, a more sophisticated model is required in this scenario.

A deformable transformation or spatially varying deformation model is able to model the local deformation and therefore used in many medical image registration tasks (see the recent review paper [34]). The deformable transformation W is defined as follow:

$$W(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}) \quad (2)$$

where each position \mathbf{x} in the image is assigned a displacement \mathbf{u} .

Besides the transformation model, a proper similarity metric and optimization scheme are essential parts in the image registration pipeline [34]. We adopt the pipeline from [25], in which the similarity metrics is defined based on modality independent neighborhood descriptor (MIND)

as follow:

$$E(\mathbf{u}) = \sum_{\Omega} |\text{MIND}_f(\mathbf{x}) - \text{MIND}_m(\mathbf{x} + \mathbf{u}(\mathbf{x}))|^2 \quad (3)$$

where Ω defines the registration region (the entire image in our case). The optimization (minimization in our case) of similarity metrics for the high-dimensional deformable transformation W is ill-posed, and regularization is generally necessary. Same as [25], we add a local regularization term to define the objective function:

$$E(\mathbf{u}) = \sum_{\Omega} |\text{MIND}_f(\mathbf{x}) - \text{MIND}_m(\mathbf{x} + \mathbf{u}(\mathbf{x}))|^2 + \alpha \|\nabla \mathbf{u}\| \quad (4)$$

This objective function is then optimized by the Gaussian-Newton method. The MIND based method is reported to be invariant to contrast differences and suitable for multi-model registration problems [25]. We empirically set weighting parameters $\alpha = 0.2$ (same as [35]) for all following experiments.

3. Experiments and result

3.1. Material

The proposed method is validated on a set of multi-modal and multi-vendor images. The dataset contains a total of 600 retinal images acquired by 5 fundus cameras (see Fig. 3), including (1) Canon Cr-1 Mark II (Canon); (2) Topcon NW300 (Topcon); (3) Nidek AFC-230 (Nidek); (4) EasyScan (i-Optics) and (5) Spectralis HRA OCT (Heidelberg) (the camera specifications are shown in Table 1). The first three are of the color fundus cameras and the last two are of the SLO cameras. The acquisition by each camera was done on the right eye of 12 healthy subjects with 5-times successive acquisitions on both the center of the fovea and optic-disc (OD), which produces 120 images (half of them fovea-centered and half of them OD centered) for each camera. The images are varied with each other slightly in terms of luminosity difference, translation and rotation. In addition, the exact region of the retina captured by each image is different. For example, for some of the fovea-centered images, the optic disc might not be completely captured. Since the Spectralis images have the smallest field-of-view (30°), they are considered as the *fixed image* for registration. The images of other four cameras (*the moving images*) are assigned to the corresponding Spectralis images (the same subject with fovea/OD centered). To be more specific, for one subject, the i -th acquired images using other cameras are only registered to the i -th acquired Spectralis image. Since the images were taken independently, it is equivalent to randomly pair these image yielding 600 (120×5) image pairs.

For processing, we removed the black region of the Canon and Topcon images (see Fig. 3) to make them into squared images as the fixed images are (see the red dashed lines). It is an automatic process and we used the image center as the square center. After that, all moving images are resized to the same size of the fixed image. These images are then used for descriptor matching and deformable registration. The cubic B-spline interpolation is used during the resize process. Because of the high contrast of the blood vessels, only the green channel of the fundus images is used for the registration where the same idea was adopted by [21]. The flowchart of our method that adopted in the experiments can be seen in Fig. 4.

3.2. Registration between color fundus images and SLO images

3.2.1. Qualitative comparison

We compared our proposed method (*method-p*) with the method proposed by Miri *et al.* [21] (*method-l*). In [21], all images are brought into the same resolution before the registration. A circular Hough transformation is first used to localize the optic discs. It results in the center and radius of the optic discs in the moving (c_f, r_f) and the fixed (c_o, r_o) images. The moving

Table 1. The detail of the fundus cameras that are used for registration. The examples shown in the last column are cropped from the original one to show the same region on one retina, where the luminosity and contrast variation among different cameras can be observed.

Camera name	Resolution	Type	Field-of-view	Number of image	Example
Canon Cr-1 Mark II	3456×2304	Color	45 °	120	
Topcon NW300	2048×1536	Color	45 °	120	
Nidek AFC-230	3744×3744	Color	45 °	120	
EasyScan	1024×1024	SLO	45 °	120	
Spectralis HRA OCT	1536×1536	SLO	30 °	120	

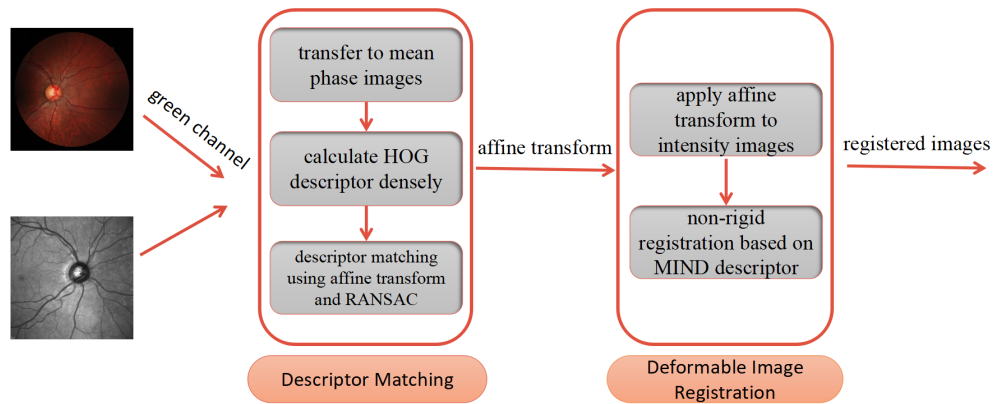


Fig. 4. Flowchart of our method.

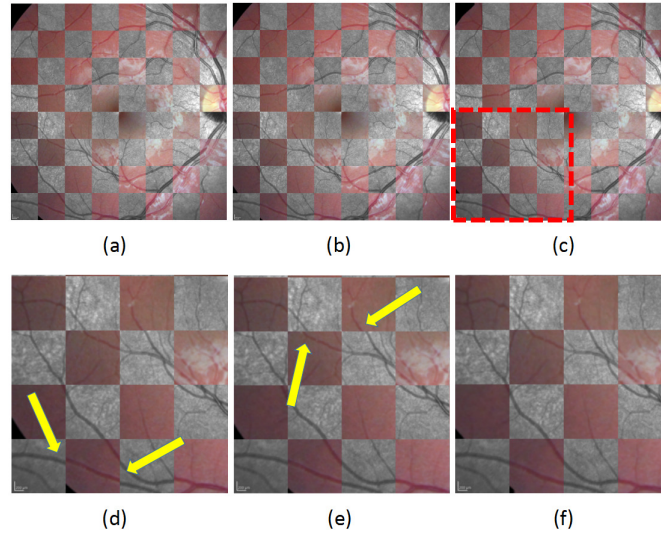


Fig. 5. Registration results between a SLO and a Canon image. (a)-(c) are from *method-1*, *method-2* and *method-p*; (d)-(f) are sub-regions from the red box of (a)-(c). The yellow arrows point out the misalignment.

images are then scaled to the fixed images by the ratio of $R = r_o/r_f$. However, this method is not applicable if the optic disc is not at the center of the image (see Fig. 2(a)). Alternatively, we calculated the rescale ratio R based on the affine transformation matrix \mathbf{A} that we derived from the descriptor matching step of *method-p*:

$$R = \sqrt{(\mathbf{A}_{1,1}^{-1})^2 + (\mathbf{A}_{2,1}^{-1})^2} \quad (5)$$

where $A_{1,1}$ and $A_{2,1}$ are the elements of the matrix \mathbf{A} .

We followed the implementation from [21] for the *method-1* (except for the disc localization). Their method was optimized for registering fundus images (Nidex and Topcon) to SD-OCT images. We used their parameter settings for our registration tasks. We also tried different parameter settings of *method-1* for our dataset but did not find any significant improvement. For comparison we also evaluated the registration performance only based on our intermediate step, the descriptor matching (*method-2*).

The registrations between Spectralis and the other four types of images are shown in Fig. 5, Fig. 6, Fig. 7, and Fig. 8. One can clearly see the misalignment (pointed by the yellow arrows) by using *method-1* and *method-2*. It proves that only using descriptor matching with an *affine* transformation is not enough. The results showed that using a deformable transformation successfully registered these images without any noticeable misalignment.

3.2.2. Quantitative comparison

In retina imaging, many systemic diseases including diabetes and hypertension are reflected by blood vessels changes such as being tortuous, narrowing and showing leakage. Blood vessels are key landmarks for inspection. To objectively evaluate the registration, we measure the matching effect between blood vessels from pairs of images.

To obtain the blood vessel segmentation, we applied the method proposed by Zhang *et al.* [36]. This technique employs a set of multi-scale Gaussian derivative filters rotated to different orientations in so-called “orientation scores”. An orientation score is a 3-D space with axis: the

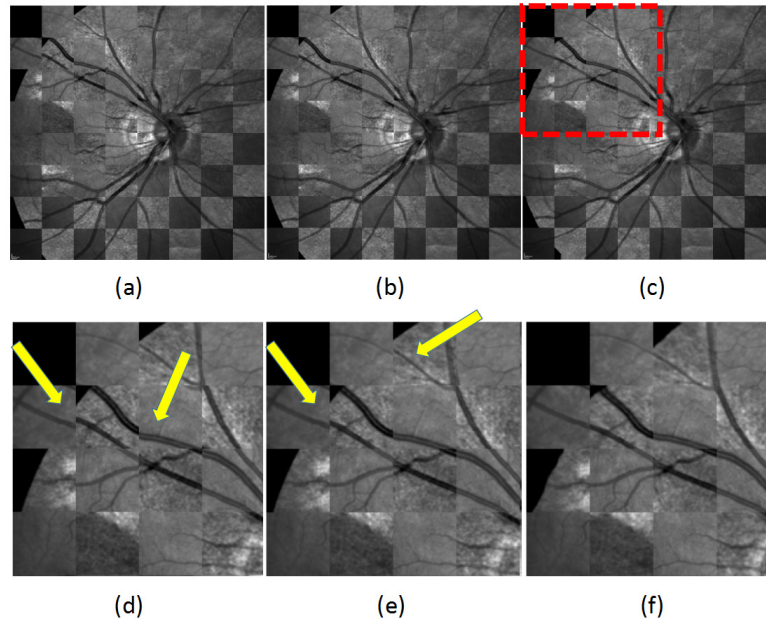


Fig. 6. Registration results between SLO and EasyScan image. (a)-(c) are from *method-1*, *method-2* and *method-p*; (d)-(f) are sub-regions from the red box of (a)-(c). The yellow arrows point out the misalignment.

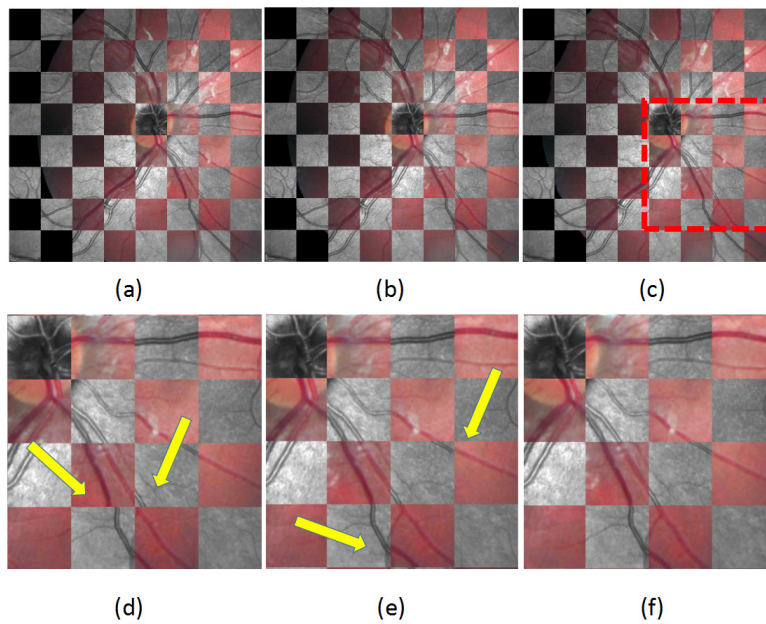


Fig. 7. Registration results between SLO and Nidek image. (a)-(c) are from *method-1*, *method-2* and *method-p*; (d)-(f) are sub-regions from the red box of (a)-(c). The yellow arrows point out the misalignment.

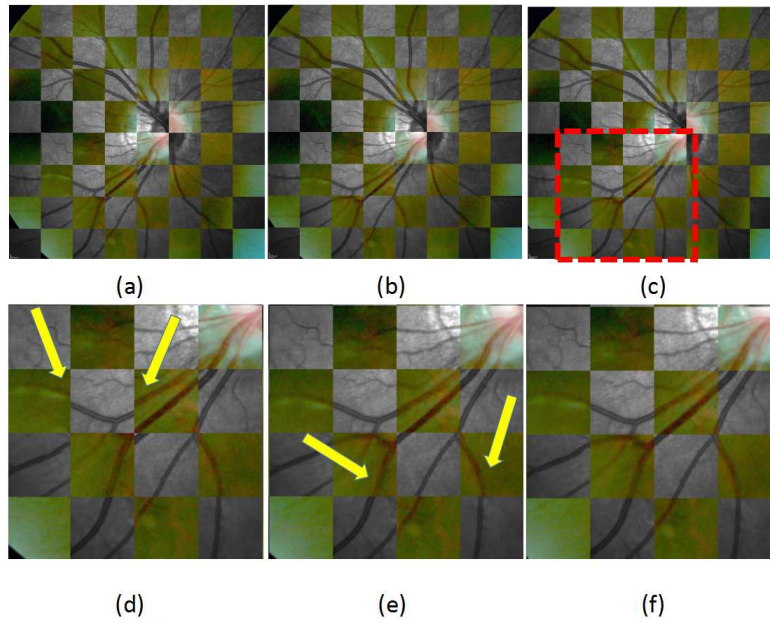


Fig. 8. Registration results between SLO and Topcon image. (a)-(c) are from *method-1*, *method-2* and *method-p*; (d)-(f) are sub-regions from the red box of (a)-(c). The yellow arrows point out the misalignment.

spatial coordinates x , y and the orientation θ , in which vessels with different orientations lay in different planes. The benefit of this construction is that difficult cases like vessel crossings are now solved because they are disentangled. The multi-scale nature of the Gaussian derivative filters ensures that disentangled vessels with various sizes are equally enhanced. Afterwards, the 3D structure is projected onto the spatial plane by taking the maximum filter response over all orientations per position. After we obtain this 2D enhanced vessel map, a proper threshold value is applied on the enhanced image to obtain a binary vascular map. Subsequently, vessels within the optic disc region are eliminated by the optic-disc mask. An iterative thinning algorithm is used to obtain the centerline of the vasculature. Junction points like vessel branchings and crossings are also removed, thus pixels connected to each other represent an individual vessel segment. In this study, we used the vessel segmentation tool which was trained on a different retinal image dataset with blood vessels manually annotated. To apply it to our 6 camera image database, we rescaled all images to the same pixel size as the training data using the size of optic disc as the reference.

To evaluate the performance of the registration methods, we calculate the Dice coefficient:

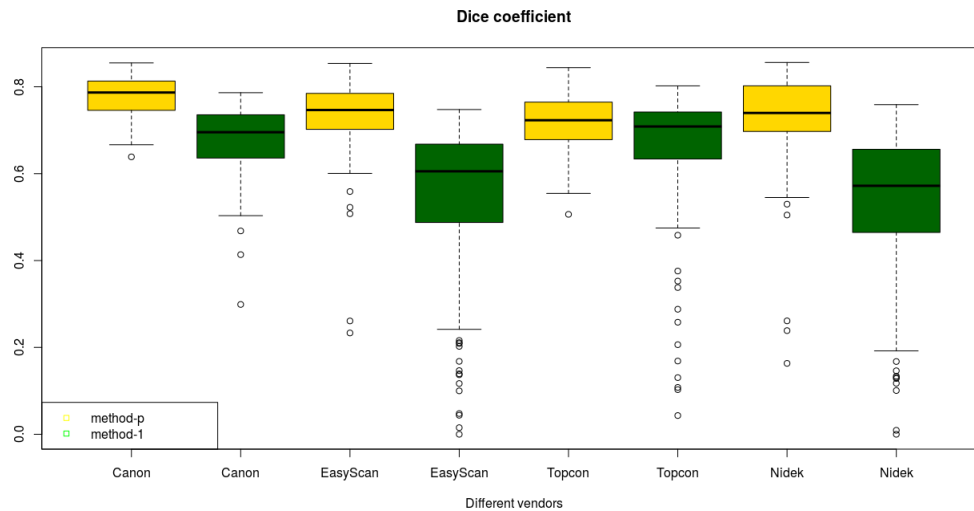
$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (6)$$

where X and Y are the vessel binary maps of fixed and moving images after registration, $|\cdot|$ represents the number of vascular pixels and $|X \cap Y|$ is the number of overlapped pixels between X and Y .

Table 2 summarizes the Dice coefficient measures from three different methods on registering four manufacturer datasets to the Spectralis dataset. Our proposed *method-p* significantly outperforms the compared methods with higher means and smaller standard-deviations (std) in

Table 2. Dice coefficient from three different registration methods

Method	Canon	Easyscan	Topcon	Nidek	All together
<i>method-1</i>	0.68 ± 0.08	0.53 ± 0.19	0.65 ± 0.16	0.53 ± 0.17	0.60 ± 0.17
<i>method-2</i>	0.66 ± 0.06	0.65 ± 0.09	0.65 ± 0.09	0.59 ± 0.09	0.62 ± 0.09
<i>method-p</i>	0.78 ± 0.05	0.73 ± 0.09	0.72 ± 0.06	0.73 ± 0.11	0.74 ± 0.08

Fig. 9. The box-plot of the Dice coefficients of our proposed method and the state-of-art method (*method-1*).

all four datasets. One can also see that *method-p* outperforms *method-1* on all cameras in terms of the means and stds. The smaller std indicates that our proposed method is more robust than *method-1*. Fig. 9 shows a box-plot of Dice coefficients of *method-p* and *method-1* on images from different cameras.

We also computed failure rates for different methods. As visually the large vessels on cases with a *Dice* less than 0.5 barely match, therefore we regards the registration is a failure. The failure rates for *method-1*, *method-2*, and *method-p* are 19.5%, 10.6% and 1% respectively. Our takes about 40s for each registration using a PC with Windows 10 64-bit OS, 32 GB RAM, and Intel(R) Core(R) CPU 4.20 GHz. For evaluation, the vessel segmentation takes about 30s per image.

4. Discussion and conclusion

In this paper, a robust and effective two-step framework is proposed to register multi-modal retinal images. The method of descriptor matching was used to register images globally in the first step. After that, a deformable registration was applied to locally refine the registration result in the second step.

In the descriptor matching process, in order to avoid the the intensity difference between different modalities, we first transferred the intensity images into mean phase images. Mean phase images are invariant to the intensity difference and often used to represent structure information of images. Since blood vessels share the same structure across different modalities, we densely calculate the HOG measure on mean phase (so called HOG-MP) image rather than only on interesting points of intensity image [21]. Points detection errors are eliminated by our

matching process. Furthermore, our registration framework does not need the vessel segmentation compared to [37, 38]. The vessel segmentation errors can also be eliminated.

We applied an affine transformation under the RANSAC framework in the descriptor matching step. However, the affine transformation is not flexible enough to model the high-order deformation between images. We used the deformation registration method based on MIND to improve the registration accuracy. MIND is also a descriptor that is invariant to modalities. It is based on the idea of self-similarity [39]. Compared to HOG-MP, the MIND descriptor however is not unique enough to satisfy the descriptor matching requirement. That is why we did not use MIND in the first step. On the other hand, the HOG-MP is not applicable in the deformation registration step due to the high computation costs.

The vessel segmentation that are used in this study is one of state-of-the-art techniques proposed in literature, which was validated on public datasets including DRIVE, STARE and CHASE DB1 giving an average of 95% accuracy. The vessel maps produced by the segmentation technique might not be similar enough, but the DICE coefficient is still able to quantify the quality of registration, as it measures the overlapping of two vessel trees instead of a single vessel. The evaluation is objective and does not require human interaction.

Our method is robust on health objects. For images obtained from pathological subjects, we would expect our method would perform similarly since findings such as hemorrhage will show up on all images. However, in the future, we will investigate the performance on images with a large time interval where findings look differently across images.

We compared our method with the state-of-art methods on registering multi-model fundus images. Both qualitative and quantitative evaluations showed that our proposed method outperforms other methods. Our multi-model images are at either fovea or optic-disc center, we haven't found any significant differences from the registration performance for both type of images. Our method generally works well if the corresponding structures (e.g. vessels) can be seen in both fixed and moving images. The evaluation is based on blood vessel segmentation across modalities. The inaccurate segmentation may influence the comparison of the Dice measures. In future work, more evaluations should be introduced. We may invite a specialist to manually select the corresponding points to set up the "golden standard" as [40] did. Furthermore, a subjective comparison may be needed [41] to evaluate the registration for the purpose of clinical application.

Funding

Netherlands Organization for Scientific Research (NWO) (629.001.003); the National Basic Research and Development Program (973 program) Program 973 (2013CB733101).

Acknowledgments

The authors would like to thank the University Eye Clinic Maastricht, Maastricht, The Netherlands for providing the retinal images of different fundus cameras. The work is part of the Hé Programme of Innovation Cooperation, which is financed by the Netherlands Organization for Scientific Research (NWO), dossier No. 629.001.003.

Disclosures

The authors declare that there are no conflicts of interest related to this article.